NAGW-937

7N-82-CR

252395

12 P

ELECTRONIC BROWSING FOR SUITABLE GIS DATA

Jeffrey L. Star, David M. Stoms, Mark A. Friedl, and John E. Estes

Remote Sensing Research Unit
Department of Geography
University of California
Santa Barbara, CA 93106 USA
(805) 961-3845

ABSTRACT

With the mass of remotely sensed data being collected, researchers are often unaware of what data is available and where it is stored. Data is frequently purchased on the basis of a terse database record describing, for instance, cloud cover and image quality. A better approach may be to allow analysts to preview spatial data, not just some salient facts about that data.

In a NASA-sponsored research program, the authors are developing a prototype system with browsing capability for data archives. The objective is to allow scientists, sitting at their local workstations, to access a network, to retrieve records of image and spatial data selected by user-specified attributes, to view low resolution versions of the data, and to place an order, if the data are satisfactory. The architecture and functioning of the BROWSE testbed is briefly outlined.

Surprisingly, the definition of BROWSE has found no consensus. This paper focuses on the on-going efforts to establish a definition that satisfies a broad range of scientific disciplines. Any operational definition will include the algorithms used to compress the raw data into a low resolution format. Preliminary algorithms being considered include single band subsetting, spatial subsampling, band ratios, principal components analysis, and linear combinations. Several browsing scenarios illustrate the complexities of selecting suitable data sets. An effective browsing utility will have benefits beyond NASA data systems and the Earth Observing System. Lessons learned from this project may be of value to other spatial data base designers.

(NASA-CR-186155) ELECTRONIC BROWSING FOR SUITABLE GIS DATA (California Univ.) 12 p
N90-70399
Unclas
00/82 0252395

1. INTRODUCTION

Effective information systems are an increasingly important aspect of spatial data analysis (Estes, 1985). Technical and social innovations provide a forcing function for improvements in the design of information systems, driven by the increasing demand for and generation of scientific data. Naisbitt (1982) discussed several social transformations that characterize corresponding changes in how geographical analysis is being done. Among these trends are shifts from:

- centralization to decentralization
- hierarchies to networking
- an industrial to an information society and
- forced technology to high technology and 'high touch', meaning increased personal involvement.

The National Aeronautics and Space Administration (NASA) has recognized these trends in establishing pilot data systems for ocean, climate, land, and planetary science data (Jet Propulsion Laboratory, 1986a; NASA Goddard, 1986; NASA, 1986; Jet Propulsion Laboratory, 1986b). Working groups are also designing information systems for the vast quantity of data projected from the Earth Observing System (EOS) on the Space Station complex in the coming decade. The challenge is twofold. The first challenge lies in developing data handling techniques for the anticipated large data volumes. The second challenge is to bring this new technology to a broader scientific and applications constituency in the service of global science (Estes et al, 1986).

The Committee on Data Management and Computation (CODMAC) of the National Academy of Science predicts that an important capability in future spatial information systems will be the ability to browse through databases from a remote terminal (CODMAC, 1982). Browsing could perform three important functions. It would allow users to locate and preview spatial data and make an informed decision about their utility, for instance, as input to a geographic information system (GIS). Secondly, it would provide a mechanism for the user community to view newly received data and make recommendations on whether the data are suitable for permanent archiving. A third function of interactive browsing would be to aid decisions regarding what additional observations to acquire during an ongoing mission (CODMAC, 1986). At NASA, this mode of operation is being called *telescience*. However, different disciplines do not necessarily share common ideas on optimal browsing formats and characteristics. This paper describes attempts to define 'browsing' in the context of multidisciplinary spatial data systems. This research is part of an on-going project at the University of California, Santa Barbara Remote Sensing Research Unit, funded by NASA Headquarters.

A brief overview of the BROWSE testbed facility at UCSB is provided. (A more complete description can be found in Star et al, 1987). Next, some tentative algorithms for generating browseable data suitable to different disciplines are proposed. In addition, several scenarios for browsing, suggested by our scientific collaborators, are outlined to suggest the complexity of reaching a consensus in a definition. Then some future research directions are discussed. The main focus of next year's effort will be in the evaluation phase. Our hope is that public discussion of these issues will not only improve the effectiveness of the BROWSE testbed, but will also stimulate creation of browsing utilities on other spatial information systems as well.

2. DESCRIPTION OF THE BROWSE TESTBED FACILITY

The first year of research, recently ended, saw the development of a rapid prototype of a browse utility at UCSB. Our objective was to get an operating version on-line and accessible via phone line to a set of collaborators from many disciplines and institutions around the United States. In the current year, we plan to incorporate their feedback on the usefulness of the system for their respective needs. What follows is a very brief description of the testbed facility.

The host facility consists of a MicroVAX II Workstation with peripheral tape drive, two 319 megabyte hard disks, a VT260 terminal, and a PC-AT as a workstation. Resident on the host computer are the database management system (DBMS), the user interface program, two databases, and a set of preprocessed browseable images. The DBMS used is a public-domain package, Relational Information Management (RIM), also being used by the NASA Ocean Data System (NODS) operated by the Jet Propulsion Laboratory.

To save the user from having to learn RIM syntax or the details of the databases, we have written a menu interface that prompts the user for query attributes. The Catalog data base contains detailed records of individual images. As seen in the Catalog menu in Figure 1, the user can restrict the search to specific geographic area, sensors, dates, maximum cloud cover, and browseable status. The graphic mode for identifying a geographic search space is shown in Figure 2, using Santa Barbara, California as an example. The program then automatically formulates the appropriate database query. The current query attributes or the current set of records retrieved can then be displayed. Figure 3 shows an example of records retrieved for the same area as shown in Figure 2. Realize that at this point in the session, only information about the data is provided, not imagery. This aspect is similar to a computerized literature search at a library or the type of query the EROS data center has supported for several years (USGS, 1980).

When a Catalog database record sounds useful, the user invokes KERMIT, a public domain communications package, to transmit the preprocessed image to their local terminal. This low resolution version can then be displayed locally using an image display program written for the IBM PC-AT workstation. Using multiple windows, the display includes a single band image, its histogram, and statistics. Currently, the display uses a level slicing routine to divide the grey values into eight colors, which can be adjusted through contrast enhancement functions in the program. Figure 4 shows a black and white reproduction of the BROWSE image corresponding to the database record of Figure 3. At this point,

```
--------------------------------------------------------------------------------
| UCSB-BROWSE      CATALOG Data Base Menu   Version 1.0  Revised 08/13/87 |
--------------------------------------------------------------------------------
|                                                                              |
|                     1. SELECT GEOGRAPHIC AREA                                |
|                     2. SELECT PLATFORM/SENSOR NAMES                          |
|                     3. SELECT DATE RANGE                                     |
|                     4. SELECT MAXIMUM CLOUD COVER                            |
|                     5. SELECT ON-LINE STATUS                                 |
|                     6. DISPLAY ATTRIBUTE CRITERIA                            |
|                     7. DISPLAY SELECTED RECORDS                              |
|                     8. DONE WITH ATTRIBUTE SELECTION                         |
|                                                                              |
--------------------------------------------------------------------------------
```

Figure 1. Catalog Menu for BROWSE
Showing Query Attributes for Restricting Search

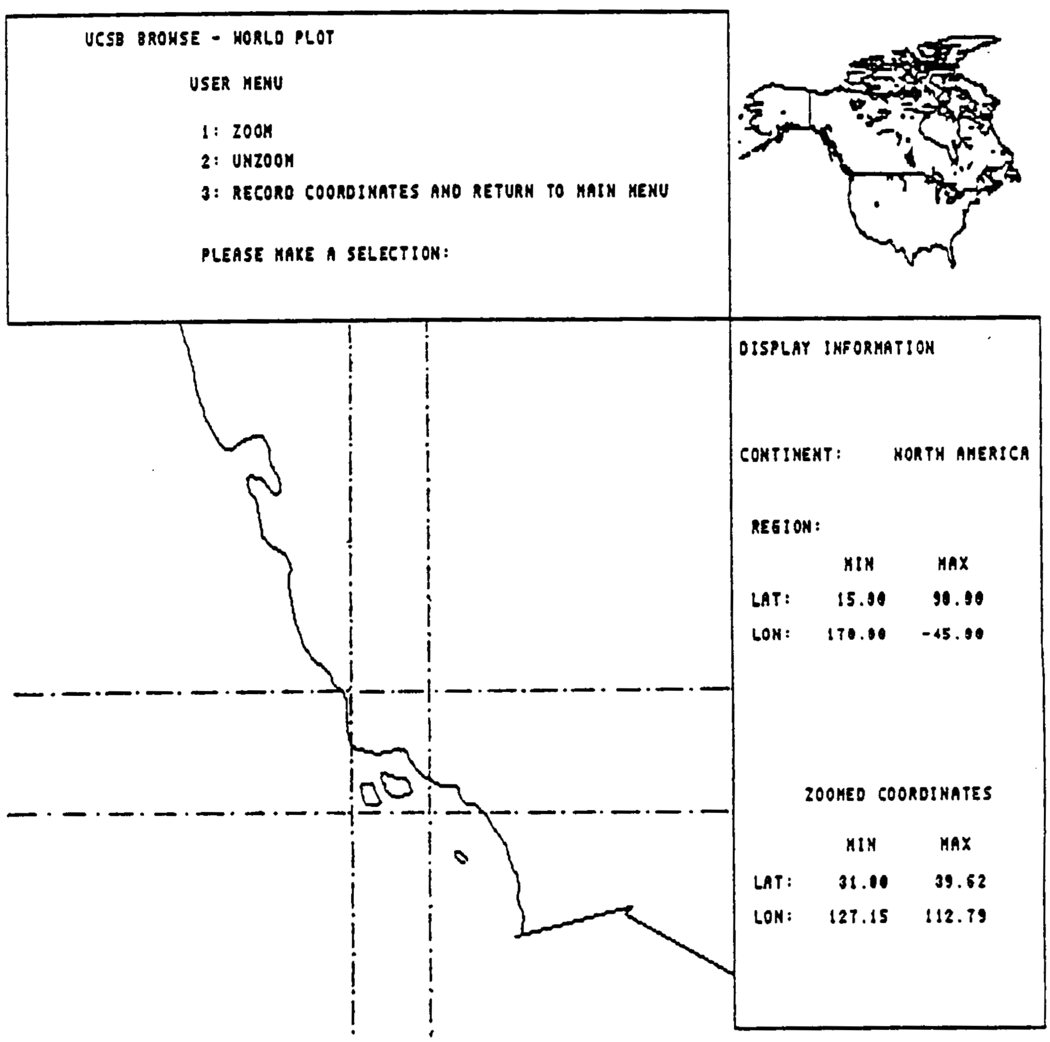

Figure 2. Example of Graphic Mode for Specifying
Geographic Region Around Santa Barbara, California

```
  UCSB-BROWSE: Current number of records selected =     3

 #    SCENE ID          CLD%    DATE     PLATFORM MISS SENSOR  ON-LINE?
---- ---------------  ----  --------  -------- ---- ------  --------

  1) Y5092817550X0PC2   20   09/15/86  LANDSAT  5    TM        ON
  2) Y5092817550X0BN3   20   09/15/86  LANDSAT  5    TM        ON
  3) Y5092817550X04/3   20   09/15/86  LANDSAT  5    TM        ON

End of current list of records

Would you like more details on any image? [Y/N]: y

Enter the number of the record you wish to see, between 1 and    3: 1

 1)   SCENE ID: Y5092817550X0PC2    DATE: 09/15/86     ON-LINE? ON
      PLATFORM: LANDSAT    MISSION: 5     SENSOR: TM
      CLOUD COVER:  20%    QUALITY: 5555555      HISTORY: PC2,SUBSET
      FORMAT: DIGITAL      PHYSICAL DESC: NULL
      FILE POINTER: [.TM]SANTAB86.PC2
```

Figure 3. Example of Data Retrieved from the Catalog
for Santa Barbara, California

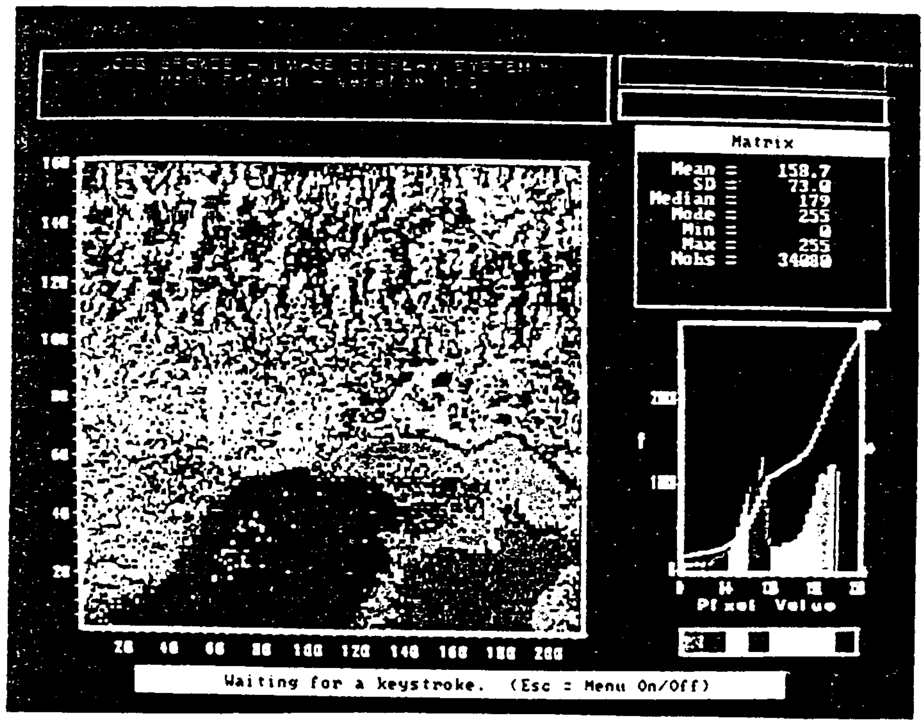

Figure 4. BROWSE Display of Scene of Santa Barbara, California
Corresponding to Record in Figure 3
(original display is in color)

the user can make a more informed decision on the suitability of the data for a spatial data processing application than from the data record alone. What makes BROWSE unique is the capability of both locating and displaying spatial data at a remote terminal.

An alternative menu path guides the user through the Directory database containing addresses and collection emphasis of various data centers around the world. Therefore, if suitable imagery is not available in the relatively small collection at the UCSB Remote Sensing Research Unit, the user can be directed to other likely sources. We plan to add Catalogs for some of these other archives to the BROWSE network. A third database of user information is planned for the coming year. This User database will permit a researcher to find other investigators who are knowledgeable about particular geographic regions, sensors, or disciplines.

3. PROPOSED BROWSING FORMATS

From the description of the testbed facility, browsing in this context can be very generally defined as locating spatial data and viewing it in some graphical form. This is not unlike how a researcher browses in a library--going to the shelf of relevant books, selecting some likely candidates, looking at the abstracts (i.e., a low resolution version of the entire text), rejecting some, finding new suggestions in the list of references, and often making serendipitous discoveries in material they were not originally aware of. It is this sense of discovery and exploration that can add Naisbitt's sense of 'high touch' to the high technology of computerized data retrieval.

Each of the NASA pilot data systems are implementing some form of browse capability. NODS includes a browse facility with data tabulations, graphics, and image formats. The user chooses a format and is presented with a list of names of browse files to view (Jet Propulsion Laboratory, 1986a). These files contain preprocessed examples of each format. Browse images of 4500 Coastal Zone Color Scanner and Advanced Very High Resolution Radiometer scenes for the West Coast Time Series have been recently put on line. At the Pilot Climate Data System, browsing includes looking at image data, but it also refers to looking at descriptions of sensors and climate parameters (NASA Goddard, 1986). Browsing or 'quick-look' support is also specified in the functional requirements for the Pilot Land Data System. Significantly, browsing is included under the data management requirements rather than those for analysis software (NASA, 1986). Recognition of this role underscores the point that browsing is a function to aid in data selection and management and should not supplant or duplicate GIS or image processing analytical functions. The Planetary Data System, is using videodisks for electronic browsing of planetary images (Jet Propulsion Laboratory, 1986b). A browse utility for EOS data is also specified (Broome, 1986).

Other data systems outside NASA are also providing browsing functions (Kubo, 1986). However, little detail of what these 'quick-look low resolution' views should show is described in any of these systems.

Among our objectives in the BROWSE project is to determine if there is a basic set of processing algorithms to generate suitable browse images. At the end of the project, we intend to provide NASA with a matrix of algorithms preferred by each discipline and suggest a small subset of these algorithms that present data in ways the majority of researchers require. The challenge is to satisfy scientists in fields as diverse as oceanography, forestry, and meteorology. The difficulty can be illustrated by considering clouds as viewed by a spaceborne multispectral scanner. To many specialists, clouds are a hindrance to information extraction. Therefore, extensive cloud cover is a negative attribute to be avoided or masked from the data. For some meteorological applications, however, it is precisely the clouds that are significant.

Providing low resolution browse data presumably involves compression in the spatial, spectral, radiometric, or temporal domains, i.e., subsampling the original data. Data compression offers several benefits in an information system and one major drawback. On the positive side, compression reduces the data storage and transmission loads on the system. Although storage capabilities are increasing (and costs decreasing) in the computer industry, the impact of EOS's proposed 262 channel High Resolution Imaging Spectrometer (HIRIS) will still be enormous in the foreseeable future (Broome, 1986). The projected data rate from HIRIS translates into one 6250 bpi density tape every second. The disadvantage of compression is that, after a certain level of reduction, the process is irreversible. Just as an abstract cannot be transformed into the original book, a highly compressed image cannot be converted by the user into the raw data. In other words, BROWSE only supplies the abstracts, as it were, so that the user can decide whether to buy the whole book.

The simplest approach to image compression is to discard all but a single spectral band. For purposes of browsing, a single band is often adequate for determining whether image data is suitable for input and analysis in a GIS. A visible wavelength band is appropriate for revealing the effects of clouds and atmospheric disturbance, for determining the presence of man-made objects, or for penetrating water surfaces. Other wavelengths have their own best uses.

A second approach capitalizes on the known relationships between spectral bands by using ratios of two bands. A ratio of reflective infrared over red wavelengths is commonly used in vegetative studies. The relative brightness is an indicator of biomass (Jackson, 1984). Geologists frequently use a mid infrared over near IR, a mid-IR over another mid-IR, or a red over blue band ratio to determine composition of surface materials (Crippen et al, in press).

Principal components analysis takes advantage of correlation in a dataset to transform it into a smaller number of dimensions. Significant scene variance may be conserved in a few synthetic output channels (i.e., components), while uncorrelated noise is collected in lower order components that can be discarded in many circumstances. We hypothesize that a single component will show enough information for a browser. The disadvantage of principal components analysis is that it can be very time consuming to compute. The transformation coefficients must be recalculated for every image, and the components may be difficult to interpret since they are scene dependent.

Linear combinations are another way to transform image data, using a predefined set of coefficients. Each combination makes a specific rotation of data in spectral space to optimize a particular feature. Perhaps the best known of this type is the Tasseled Cap transformations for Landsat Multispectral Scanner data emphasizing soil brightness and vegetation greenness (Kauth et al, 1976). A similar transformation has been developed for Thematic Mapper data (Crist et al, 1984). In essence, the first transformed axis is an ordination from wet to dry soil, while the second axis indicates biomass as a distance from the bare soil baseline. Coefficients can be easily derived for other sensors but the interpretation of the axes may be unique for each type (Jackson, 1983).

4. SOME BROWSING SCENARIOS

It is impossible to predict all the ways researchers will want to use the BROWSE facility. The collaborators who have assisted us by describing their browsing needs tend to be experienced in remote sensing applications. The new constituency NASA hopes to reach is harder to identify, and new users may have difficulty articulating what they need, other than that the system be helpful and flexible. This section describes a few of the scenarios which our multidisciplinary collaborators have identified as the process they might use in gathering GIS data.

As we progress into the age of global science, researchers are collecting datasets for large regions at high resolution. An analyst working at this scale might need to see a low resolution, single band view of many adjacent scenes acquired in a particular season by the same sensor under relatively clear atmospheric conditions.

Research projects sometimes involve multistage sampling, with low resolution data of large areas and higher resolution for selected samples. A browse facility could determine if data of all required types were available and suitable at the study site. Alternatively, when multisensor data is required, the DBMS could identify sites with all the required types. By browsing through these sites, the analyst could select the most suitable for further investigation and commitment of staff and funds.

Discipline specialists new to remote sensing may not know the capabilities of different sensors. In an extreme example, an investigator wanting to census elk populations over a large region may believe that a single scene of a low resolution sensor like the Advanced Very High Resolution Radiometer (AVHRR) would be adequate. A quick-look at a browse image should convince them immediately that AVHRR data in not appropriate for the task.

Misplacing research data over time as students come and go is an embarrassing fact of life in many academic departments. It is hopeless, as we ourselves have discovered, to go through a storeroom full of poorly labeled tapes, trying to find a valuable dataset someone in the past has painstakingly assembled. If a browse image of each dataset were available, along with a trail to the dataset itself, researchers might spend less time recreating data.

5. FUTURE DEVELOPMENT PLANS

The first year of development focused on rapid prototyping of the BROWSE testbed facility and exploring what 'browse' means to different disciplines. The existing prototype consists of a single node at UCSB with a relatively small database. In the coming grant year, we intend to expand the system to include at least one or two additional nodes with different types of image data. To accomplish this involves developing a high level activity manager that queries the appropriate databases around the network, transparently to the user, so a person does not need to learn multiple query languages.

The prototype facility has now become accessible via standard phone lines to our collaborators. We will be monitoring their use and incorporating their feedback on the user interface and the types of browse formats. In concert with this, we will be examining data compression in greater detail. Tradeoffs between storage overhead of preprocessed image data versus time delays of processing on demand will be assessed.

The existing prototype assumes the user can select appropriate sensors and will recognize when data are suitable. Since NASA hopes to establish a wider constituency for data acquired in space, more expert guidance will be necessary to aid new users in formulating appropriate queries. Development of an expert system is planned that can give intelligent advice on query attributes. The user need respond only with the parameters of their task i.e., the view that is more familiar to them. The initial expert enhancement will be in the domain of vegetation applications where a substantial knowledge base can be assembled from the research literature.

6. CONCLUSIONS

A capability to browse spatial data will be an important function in future spatial data systems. Researchers will browse in order to locate suitable input data, to recommend imagery for permanent archiving, and to request additional data acquisitions from ongoing missions. When EOS begins generating massive volumes of data in support of global science, the need for these three functions will be even greater. The BROWSE project outlined here indicates one effort by NASA headquarters to prepare for the EOS era.

As defined above, 'browsing' involves both locating and viewing spatial data. The browse function is most appropriate as a data management tool. We do not see it as a replication of existing GIS or image processing analytical functions. Several methods of compressing data into browseable format have been discussed. Additional methods, from disciplines different from those of the authors, may be identified during the testing and evaluation phase. Planned enhancements should make BROWSE more effective in serving a wider constituency of users, which is one of NASA's objectives. To be effective, a browsing function should also be fast, flexible, and friendly.

Persons interested in joining the list of BROWSE users/testers are encouraged to contact the authors at the above address. Together we can give browsing for spatial data a degree of 'high touch' comparable to the level of high tech provided by computerized database management.

Acknowledgments

The authors would like to acknowledge the contributions of Dr. Tony Villasenor and Dr. Jim Dodge at NASA Headquarters, Dr. Chuck Klose, Pat Hogan, and Iveory Johnson at the Jet Propulsion Laboratory, and those numerous individuals who are providing us with much useful input and data. Research for this project was sponsored under NASA grant NAGW-987.

References

Broome, Jr., Douglas R., "The Earth Observing System: A Multidisciplinary System for the Long-Term Acquisition of Earth Science Data from Space," *Monitoring Earth's Ocean, Land, and Atmosphere from Space--Sensors, Systems, and Applications*, vol. 97, pp. 797-822, 1986.

CODMAC, *Data Management and Computation*, 1: Issues and Recommendations, 167 pp., National Academy Press, Washington, D.C., 1982.

CODMAC, *Issues and Recommendations Associated with Distributed Computation and Data Management Systems for the Space Sciences*, 111 pp., National Academy Press, Washington, D.C., 1986.

Crippen, R.E., R.G. Blom, and J.H. Heyada, "Directed Band Ratioing for the Retention of Perceptually-Independent Topographic Expression in Chromaticity-Enhanced Imagery," *International Journal of Remote Sensing*, in press.

Crist, Eric P. and Richard C. Cicone, ''A Physically-Based Transformation of Thematic Mapper Data--The TM Tasseled Cap,'' *IEEE Transactions on Geoscience and Remote Sensing*, vol. GE-22, no. 3, pp. 256-263, 1984.

Estes, John E., ''The Need for Improved Information Systems,'' *Canadian Journal of Remote Sensing*, vol. 11, no. 2, pp. 124-131, 1985.

Estes, J. E. and J. L. Star, ''Support for Global Science: Remote Sensing's Challenge,'' *IEEE Proceedings*, vol. 73, no. 6, pp. 1097-1107, 1986.

Jackson, Ray D., ''Spectral Indices in n-Space,'' *Remote Sensing of Environment*, vol. 13, pp. 409-421, 1983.

Jackson, Ray D., ''Remote Sensing of Vegetation Characteristics for Farm Management,'' *Proceedings of SPIE*, vol. 475, pp. 81-96, 1984.

Jet Propulsion Laboratory, *Software Specification for the GOLD Catalog*, 74 pp., California Institute of Technology, Pasadena, CA, 1986a.

Jet Propulsion Laboratory, *Planetary Data System Version 1.0 System Specification*, 234 pp., California Institute of Technology, Pasadena, CA, 1986b.

Kauth, R.J. and G.S.Thomas, ''The Tasseled Cap--A Graphic Description of the Spectral-Temporal Development of Agricultural Crops as Seen By Landsat,'' *Symposium on Machine Processing of Remotely Sensed Data*, pp. 4B-41 through 4B-51, 1976.

Kubo, Sachio, ''The Basic Scheme of TRINITY: A GIS with Intelligence,'' *Proceedings of the Second International Symposium on Spatial Handling*, pp. 363-374, Seattle, WA, 1986.

Naisbitt, J., *Megatrends*, 333 pp., Warner Books, New York, NY, 1982.

NASA, *Pilot Land Data System Functional Requirements--Version 0.2*, 13 pp., , 1986.

NASA Goddard Space Flight Center, *Proceedings of the Second Pilot Climate Data System Workshop, NASA Conference Publication 2430*, 263 pp., National Space Science Data Center, Goddard Space Flight Center, Greenbelt, MD, 1986.

Star, Jeffrey L., Mark A. Friedl, David M. Stoms, and John E. Estes, ''Browse Capability for Spatial Databases,'' *Proceedings of GIS'87 Conference*, San Francisco, CA, 1987.

USGS, *A User's Guide to the Teletype-Inquiry Option of the INORAC System*, EROS Data Center, Sioux Falls, SD, 1980.

Biographical Sketches

Dr. Jeffrey L. Star is manager of the Remote Sensing Research Unit and Lecturer at the University of California, Santa Barbara. He received his undergraduate degree from M.I.T. and his doctorate in oceanography from Scripps Institution of Oceanography, University of California, San Diego. His research interests are in remote access to large archives of spatial data, merging artificial intelligence with geoprocessing technologies, and creation of a regional GIS for agricultural monitoring. Dr. Star has served on a number of national committees, including NASA's Earth Observing System Data Panel and the Task Force on Geo Information of the Research Libraries Group.

David M. Stoms is a Ph.D. student in the Department of Geography at the University of California, Santa Barbara. His Masters thesis was entitled 'Preliminary Design of a Farm Monitoring Geographic Information System'. Prior to graduate school, he worked 12 years with the U.S. Forest Service, primarily in land management planning. His current research interests are in browsing capabilities for spatial databases, knowledge-based approaches to spatial data processing and retrieval, and the value of improved spatial information.

Mark A. Friedl is a Masters student in the Department of Geography at the University of California, Santa Barbara. He received his B.S. from McGill University in Montreal in geography in 1986. While at McGill, he was involved with research at the McGill Subarctic Research Station and with the McGill Advanced Cartography Laboratory. He is currently involved in research related to spatial data handling and knowledge-based interpretation of remotely sensed imagery.

Dr. John E. Estes is Professor of Geography and Director of the Remote Sensing Research Unit at the University of California, Santa Barbara. He has authored more than 250 works on remote sensing and information systems, including 'Fundamentals of Image Analysis', a chapter in the *Manual of Remote Sensing*, which he coedited. He has served as senior scientist at the National Aeronautics and Space Administration and is a member of the Committee on Data Management and Computation of the National Academy of Science.